IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ryan Smith
11-6-2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Using data sourced from a combination of SpaceX directly via their SpaceX API and Wikipedia via web scrapping, this report will give key insights into the success of the Falcon 9 rockets with their reusable stage one.

The data presented in this report was cleaned and pre-processed using the Python Numpy and Pandas libraries to remove all null values, organize the launches by launching site and convert the mission success data into binary.

The exploratory analysis was done using the Python Seaborn, Folium and Dash librabries to begin to answer the question of how much effect aspects of the launches such as site location, orbit followed and payload had on the success or failure of the missions

The findings indicate that the Kennedy Space Center recorded the top successful landing rate (76.9%). Payloads in the range of 2,000 to 4,000 kilograms showed the steadiest success in landings. The FT booster rocket proved to be the most dependable type overall. When it comes to the predictive models that were built, a Decision Tree Classifier stood out after hyperparameter tuning. It reached the best training accuracy at 88.8 percent. Analysis of the confusion matrices suggests that mistakes in these models tended toward false positives. False negatives appeared less common.

Introduction

SpaceX has emerged as a leader in the cost-efficient commercialization of space travel, with its Falcon 9 launches costing approximately $62 million USD—significantly lower than the $165 million USD launches of competing platforms. This financial advantage is largely driven by the company's ability to reuse its first-stage rockets, a breakthrough that dramatically reduces launch costs. As new competitors enter the industry, achieving comparable or greater reusability success will be essential to remain competitive.

This project explores the factors contributing to SpaceX's successful first-stage recoveries and applies machine learning techniques to predict future landing outcomes. Using historical launch data, the analysis demonstrates that predictive models can forecast the success or failure of Falcon 9 landings with nearly 90% accuracy, offering valuable insights into the drivers of reusable rocket reliability.

Section 1

Methodology

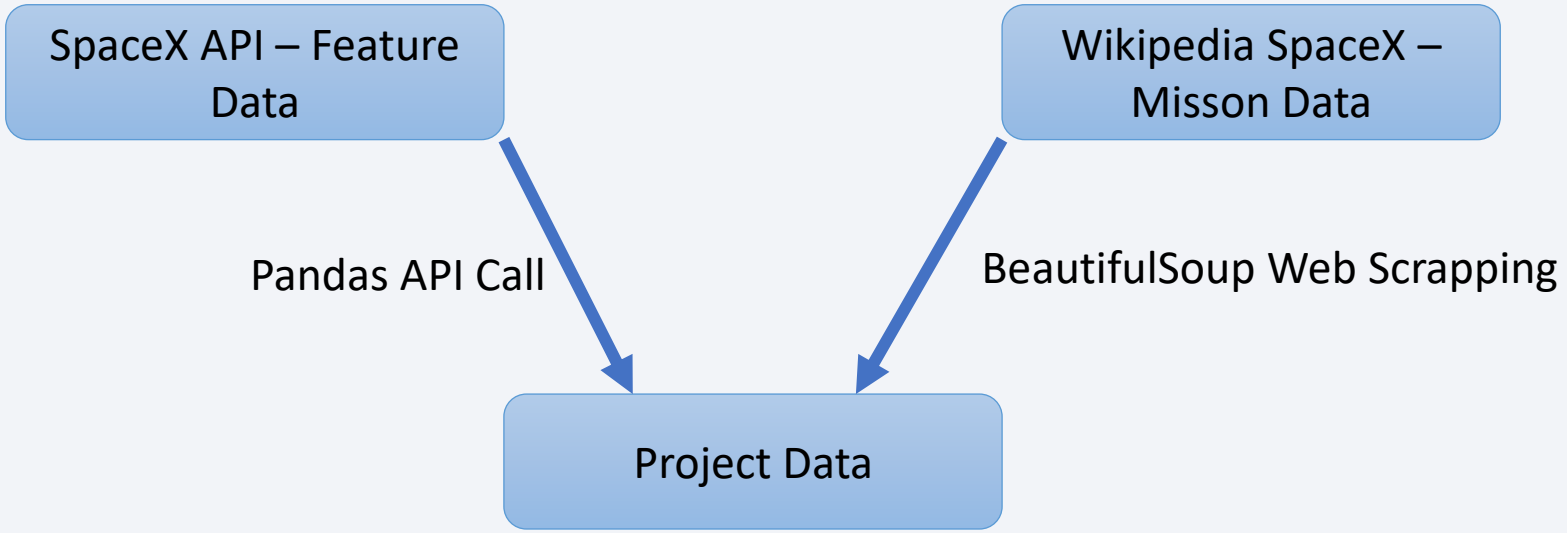
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - Wikipedia – Via web scrapping with BeautifulSoup Python library
- Perform data wrangling
  - Removed null values
  - Created a binary target feature 0=mission failure, 1=mission success
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Split the data into training and testing sets
  - Trained four models with optimized hyper parameter and 10-fold cross validation

Data Collection

- SpaceX offers an API to directly acquire some of their data. For this data the API was called using Pandas to pull the data into a data frame.
- For the data which was not available directly from their API, the data was pulled from Wikipedia via web scrapping with the BeautifulSoup Library.

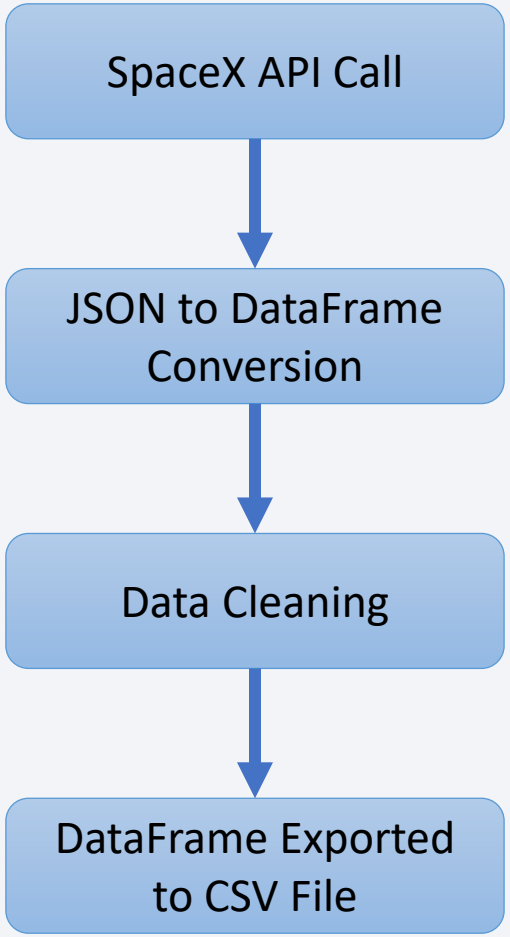


Data Collection – SpaceX API

- SpaceX API connection was made using Python requests.
- The response JSON was then converted to a Pandas DataFrame using json_normalize().
- The DataFrame was then filter to remove irrelevant features and *null* values.
- Lastly the DataFrame was converted and exported to a CSV file using to_csv().

Jupyter Notebook

CSV File

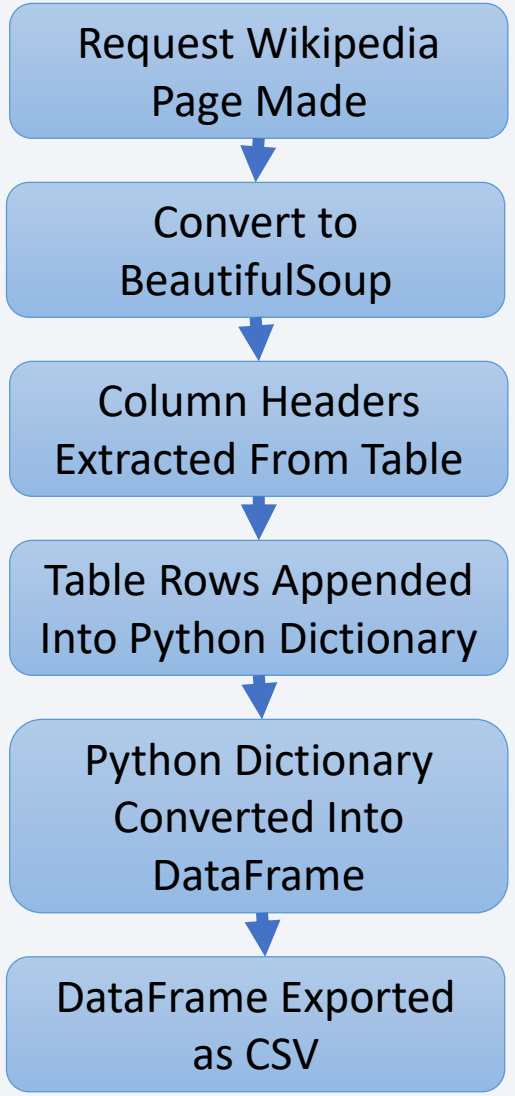


Data Collection - Scraping

- Python get request was made to SpaceX's Falcon 9/Falcon Heavy Launch Wikipedia page.
- The response was converted to a BeautifulSoup object with header name extracted form the table headers.
- The table rows were then iteratively appended to into a Python dictionary.
- The python dictionary was then converted to a DataFrame
- Lastly the DataFrame was converted and exported as a CSV file.

Jupyter Notebook

CSV File

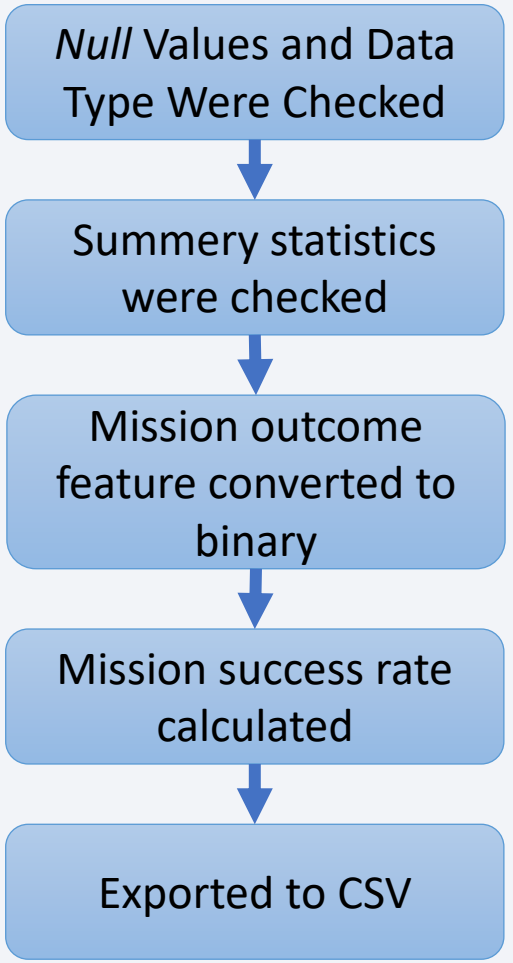


Data Wrangling

- Data were check for *null* values and correct data types.
- Summery data were checked in relevant fields, namely launch site, mission outcome and orbit.
- The mission outcome feature was converted into a binary value.
- The overall mission success rate was found to be approximately 67%.

Jupyter Notebook

CSV File



EDA with Data Visualization

- Scatter plot of launch number versus launch site color coded by mission outcome was made looking for correlation between variables.
- Scatter plot of payload versus launch site color coded by mission outcome was made also looking for correlation.
- Bar plot of mission outcome for each orbit value was made to identify the orbit with the best outcome rate.
- Scatter plot of Launch Number versus orbit color coded by mission outcome was made looking for correlation.
- Line plot demonstrating mission outcome over time was made showing a strong increase in mission success over time.

Jupyter Notebook

EDA with SQL

SQL queries were made to find

- Names of all launch sites.
- Data from five launches where the site began with “CCA”.
- Total mass of payloads carried by NASA.
- Average payload mass carried by F9 v1.1 boosters.
- Date of first successful mission outcome on a ground pad.
- Boosters which have had positive mission outcomes landing on a drone ship with payload between 4,000 and 6,000 Kg.
- Total Number of each mission outcome.
- Which boosters have carried the maximum payload.
- Missions which have failed landing on drone ship in 2015.
- Ranked count of mission outcomes between June 4, 2010 and March 3, 2017.

Jupyter Notebook

Build an Interactive Map with Folium

- Markers were placed on the map of all launch sites with the purpose of examining surrounding geographical features.
- At each launch site color coded cluster markers were placed for each mission. Green for successes and red for failures. This visually displays the successes and failures of the launch sites.
- Lines were place between a launch site and the nearest coast to examine the distance.

Jupyter Notebook

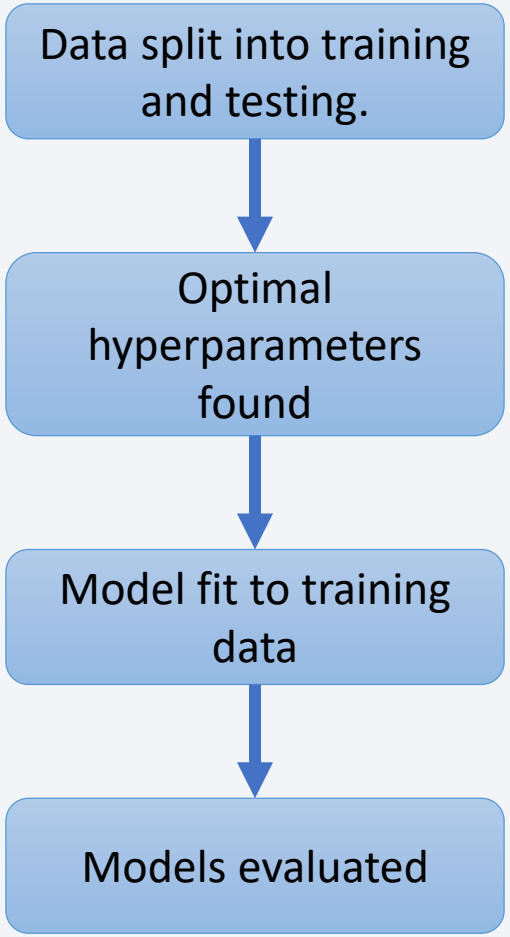
Build a Dashboard with Plotly Dash

- An interactive dashboard containing a launch site drop down and payload slider was created.
- This dashboard displays a pie chart demonstrating the overall success of launch sites,
- A scatterplot visualizing mission outcome for the launch sites and payload carried.

Python Code

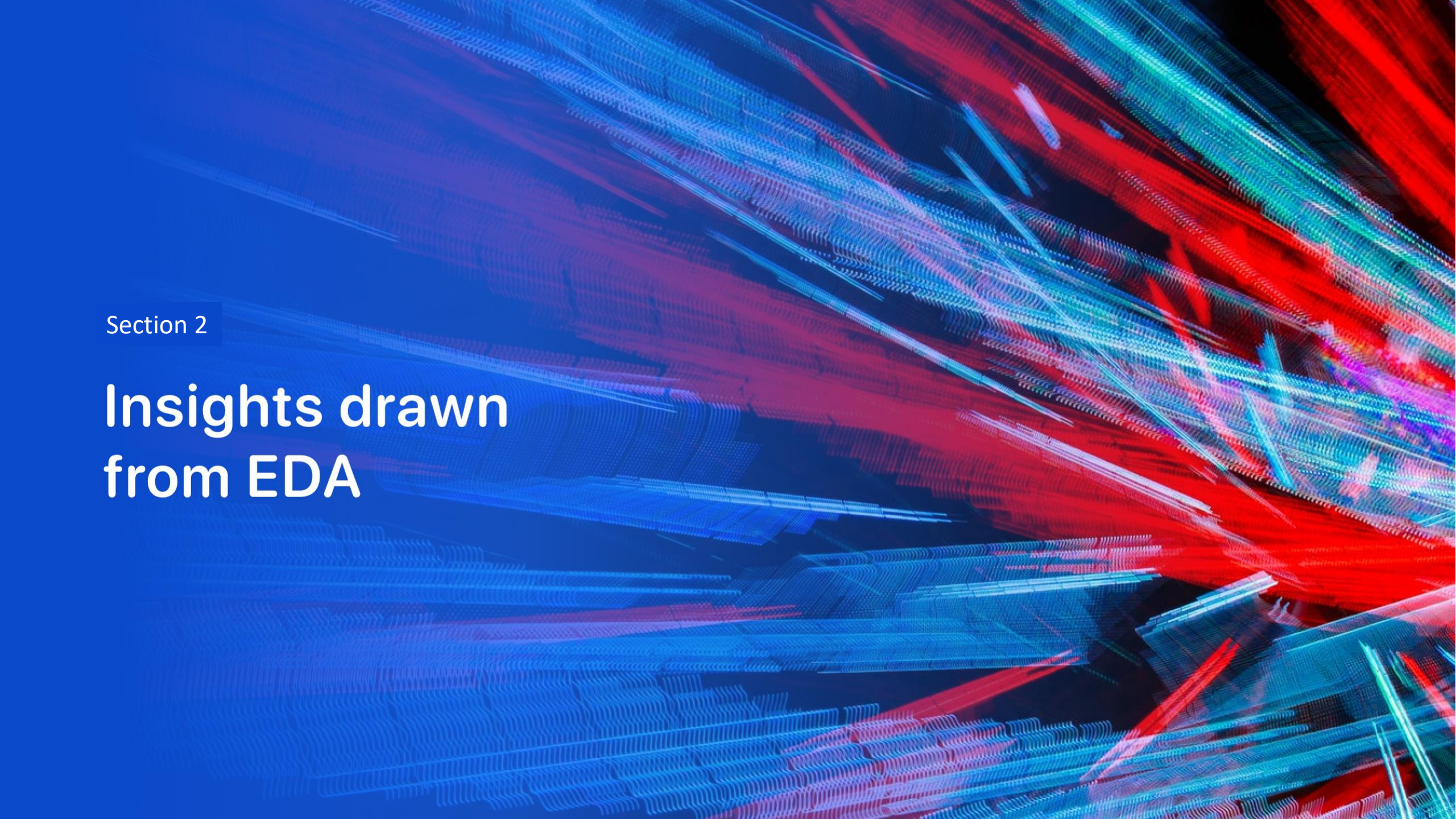
Predictive Analysis (Classification)

- Data was split into training and testing groups.
- GridSearchCV was used to determine optimal hyperparameters for four different models.
- The models were fit with training data using 10-fold cross validation.
- The models were then tested for accuracy on the testing data and a confusion matrix was created.

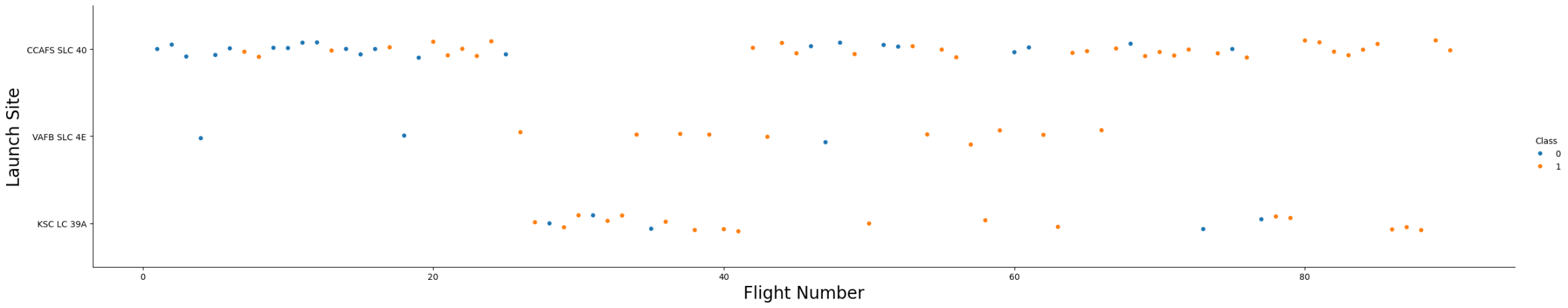


Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

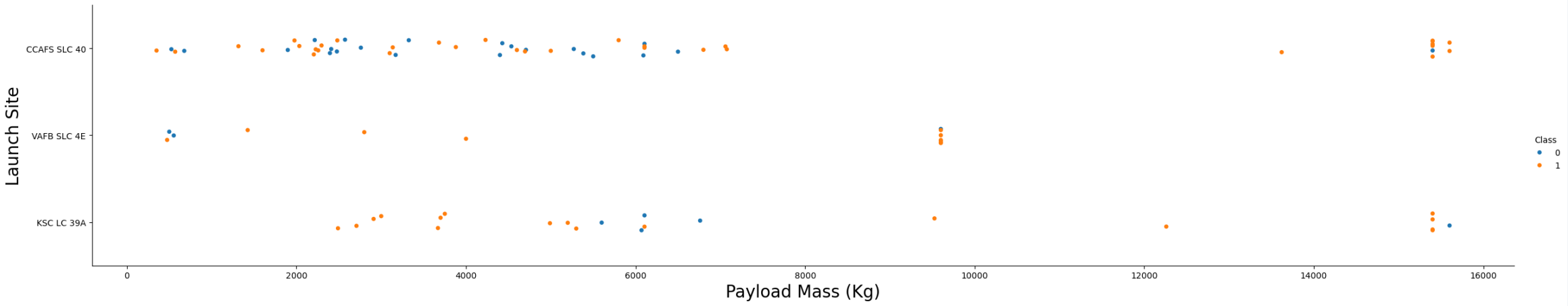
Flight Number vs. Launch Site



This visualization shows that the early years of Falcon 9 launches were failure prone and launched nearly exclusively from Cape Canaveral.

Kennedy Space Center has the best success rate among the launch sites and started launching with SpaceX well into the Falcon 9 program.

Payload vs. Launch Site



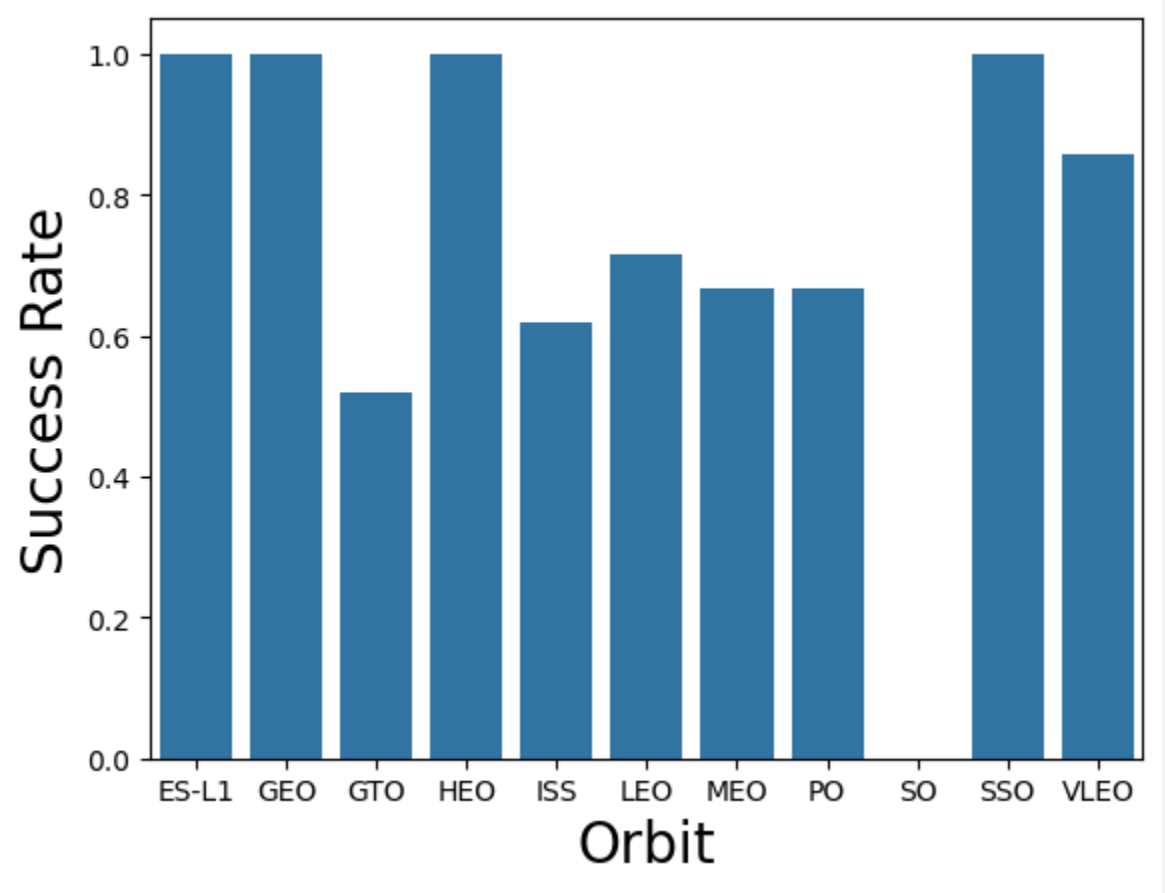
Cape Canaveral and Kennedy Space Center have launched carrying payloads greater than 10,000 Kg with nearly all being successes.

SpaceX generally launch with payloads less than 8,000 Kg from all sites and mission outcome is much more mixed than larger payloads.

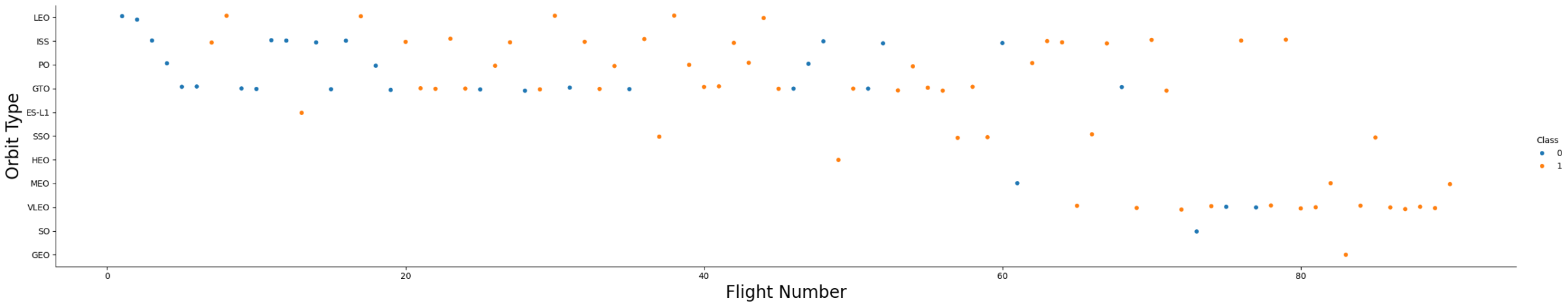
Success Rate vs. Orbit Type

While several of the orbit types have a 100% success rate there are very few instances of launches to those orbits.

The highest success rate for frequently used orbit types (>10) is VLEO at over 85%.



Flight Number vs. Orbit Type



In the early years of SpaceX there is a cluster of high failure rate launches as is expected with innovation.

The VLEO orbit type is relatively new to SpaceX while holding the best success rate among orbits as found earlier.

Payload vs. Orbit Type

Heavy payloads are generally more successful than lighter payloads overall.

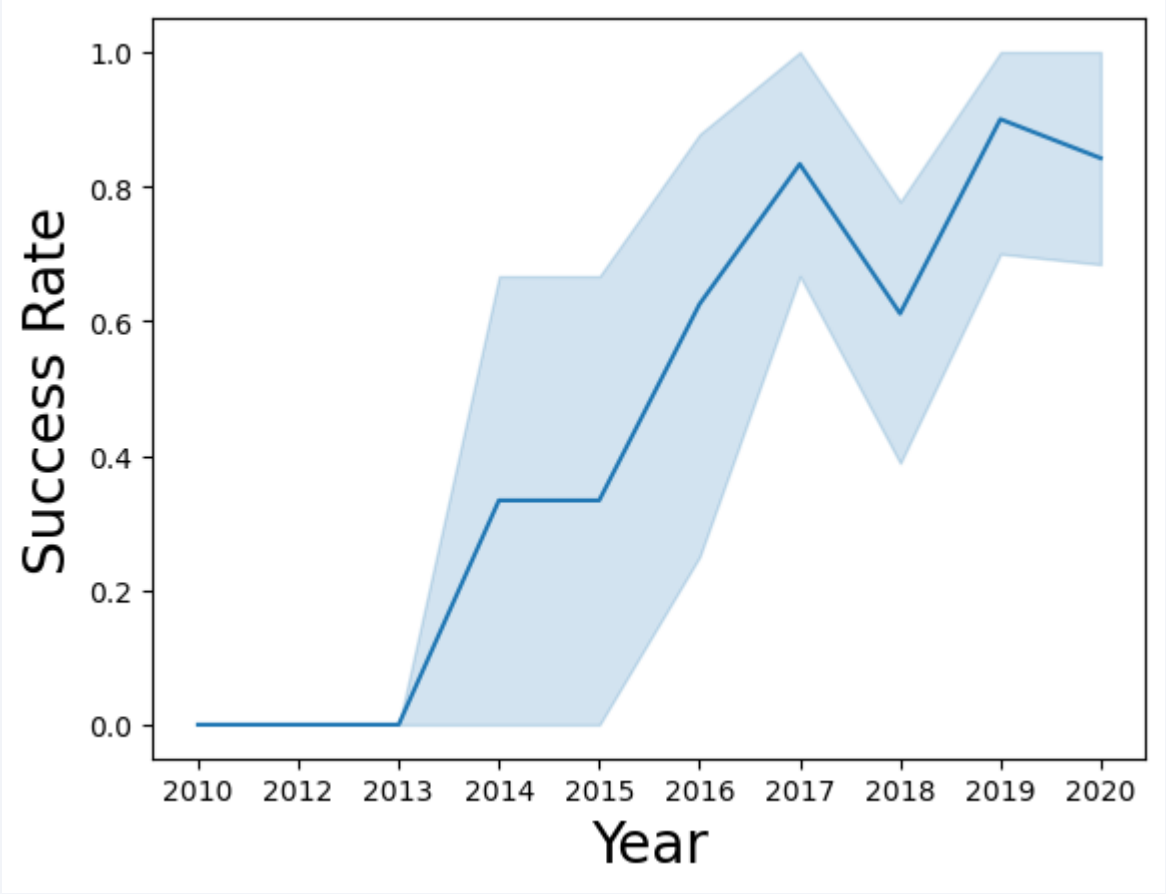
The VLEO orbit type is used to carry the heaviest payloads among the orbits.

Launch Success Yearly Trend

The early years of failures can clearly be seen

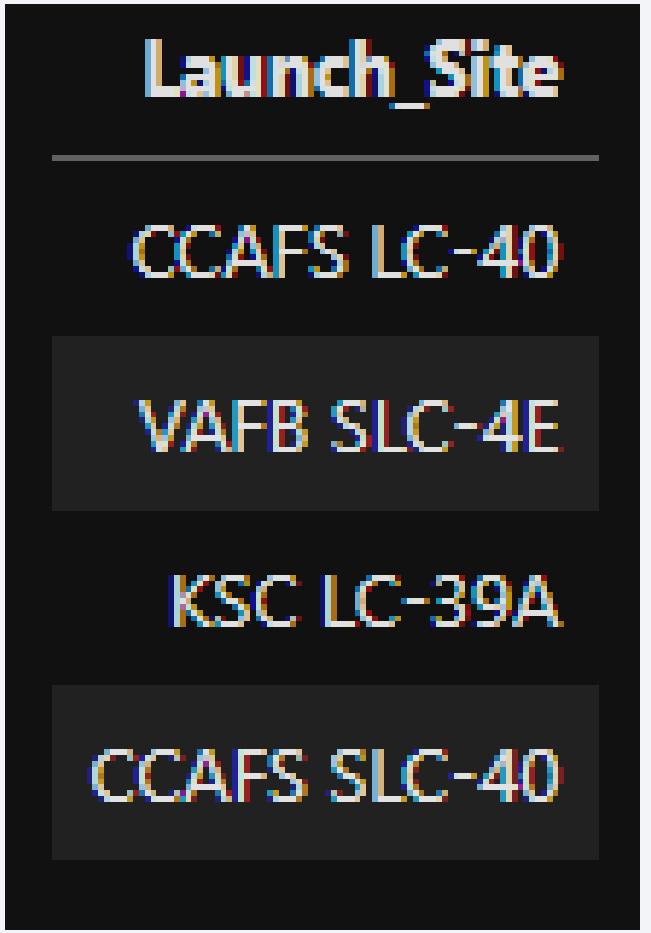
SpaceX has been steadily increasing their success rate on average since 2013.

The peak of success was in 2019 at approximately 90%



All Launch Site Names

All Launch site names pulled with a SQL query



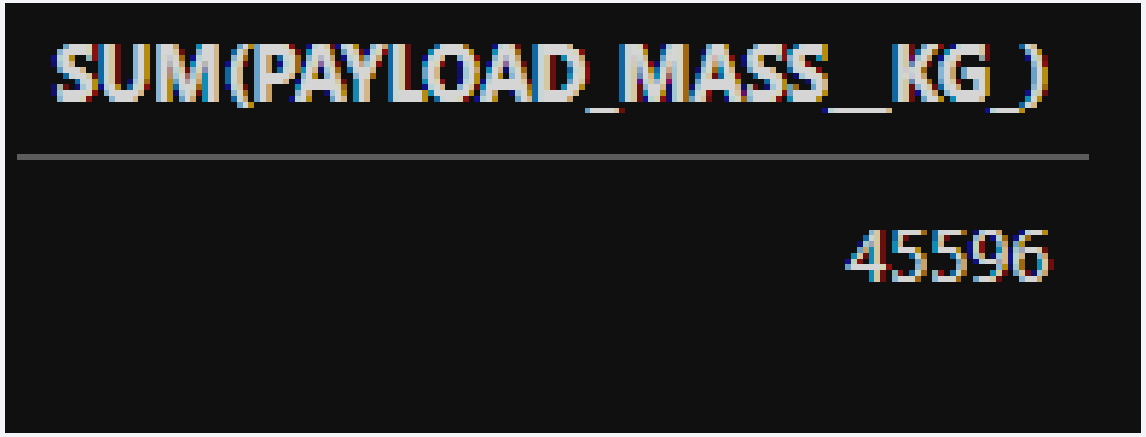
Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_(
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p:
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (p:
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	Nc
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	Nc
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	Nc

The first five results from the SQL query of the dataset.

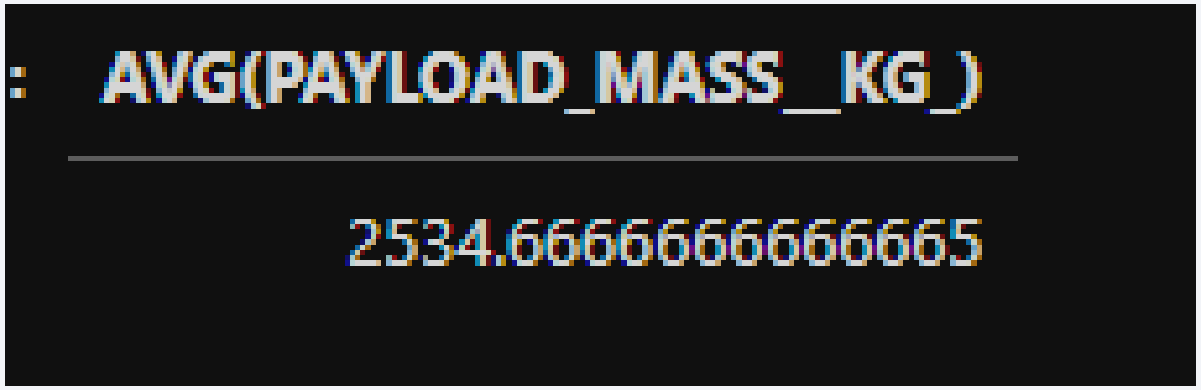
Total Payload Mass

Total payload mass as calculated by SQL query 45596 Kg.



Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1 was 2534.7 Kg



First Successful Ground Landing Date

The date of the first successful mission landing on a ground pad was December 22nd 2015

MIN("Date")
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of the boosters which carried between 4,000 and 6,000 Kg while successfully landing on a drone ship.

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Table showing the number of successful and failed mission outcomes.

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Table showing the names of boosters which have carried the highest payload mass (15,600 Kg)

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

Table showing the list of failed landings on a drone ship in 2015.

SUBSTR("Date",6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Ranking of the count of mission outcomes between June 4th 2010 and March 3rd 2017. Shown in descending order.

Landing_Outcome	COUNT(*)
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

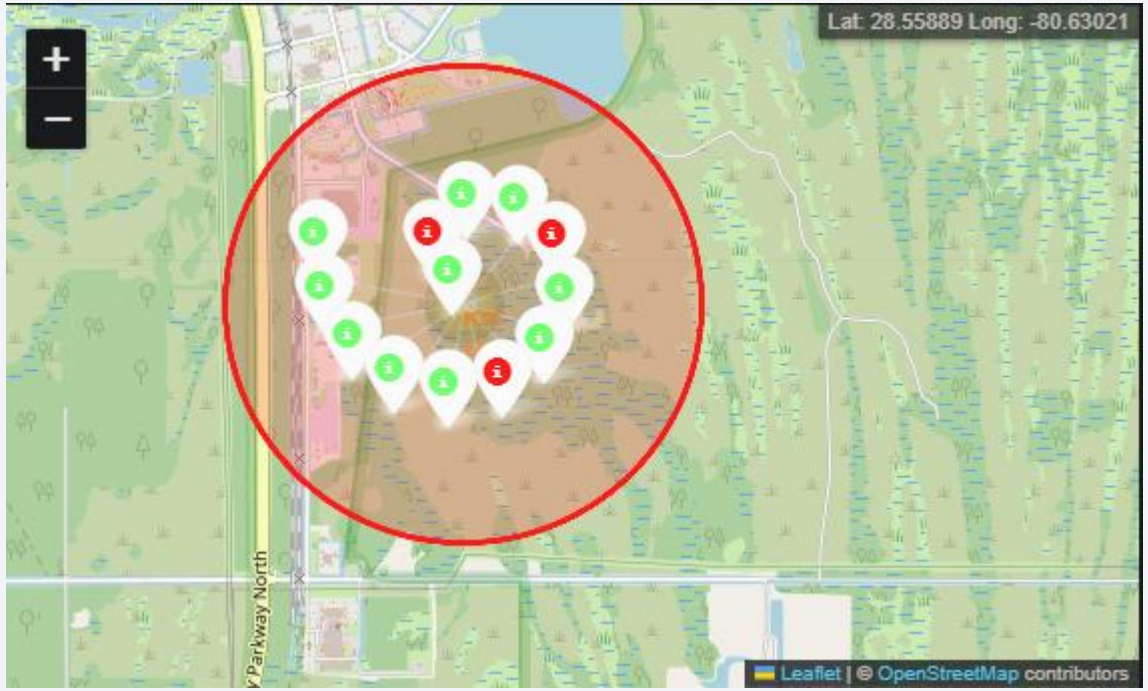
Launch Sites Proximities Analysis

Map of All SpaceX Launch Sites

Map showing the launch site used by SpaceX are cluster along the coasts of the United States of America.

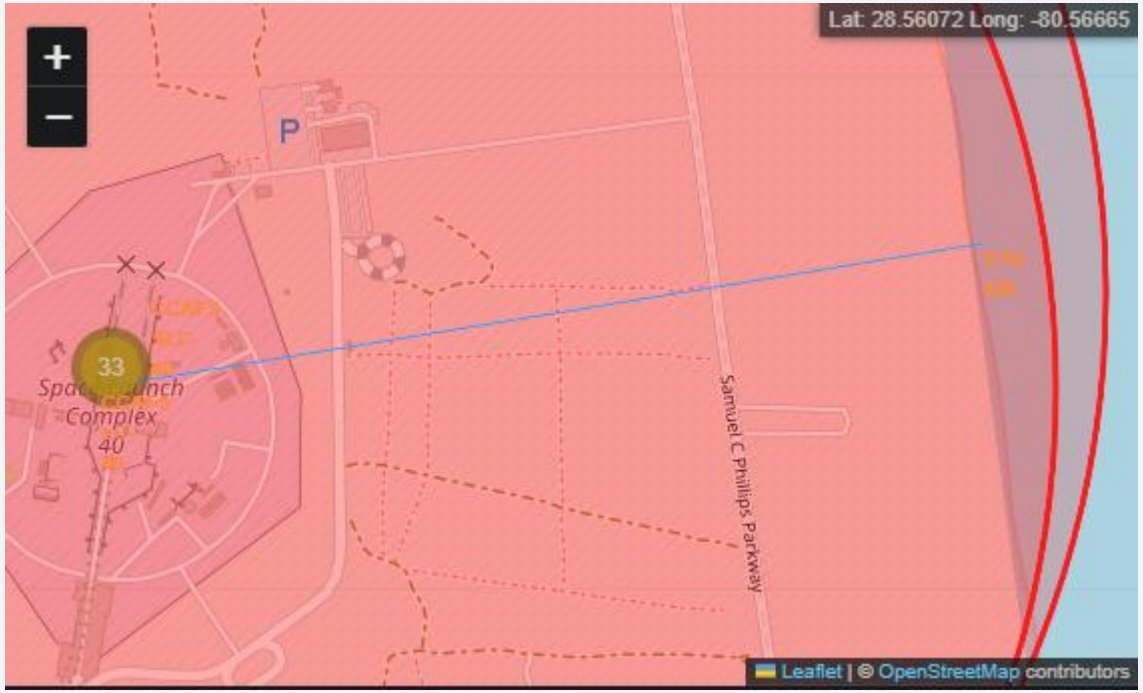
Mission Outcomes of Site KSC LC-39A

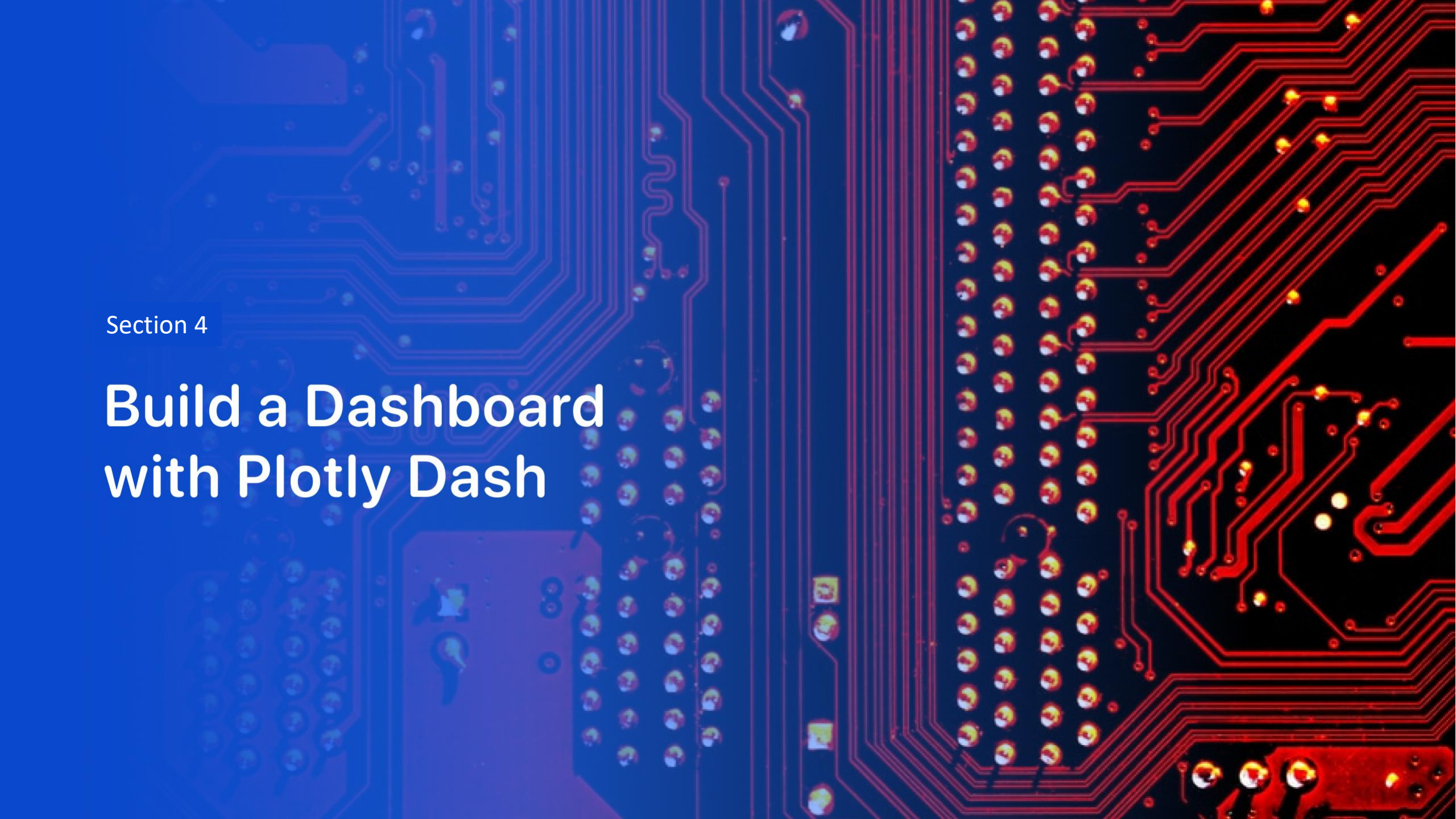
Map shows the mission outcomes for the KSC LC-39A cluster.



Falcon 9 Launch Pads Near Coastlines

Map showing the distance from the Falcon 9 launchpad to the nearest coast (0.94 KM)

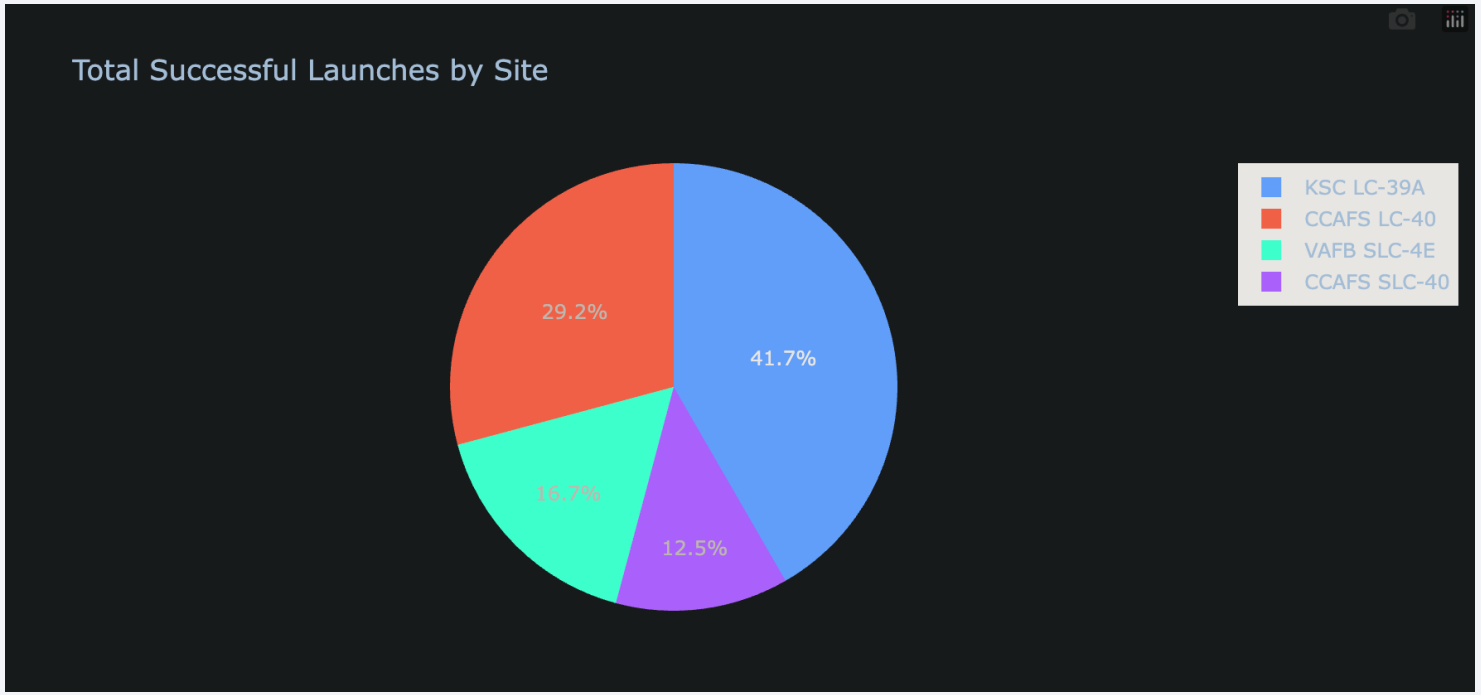



Section 4
Build a Dashboard with Plotly Dash

SpaceX Total Launch Successes by Site

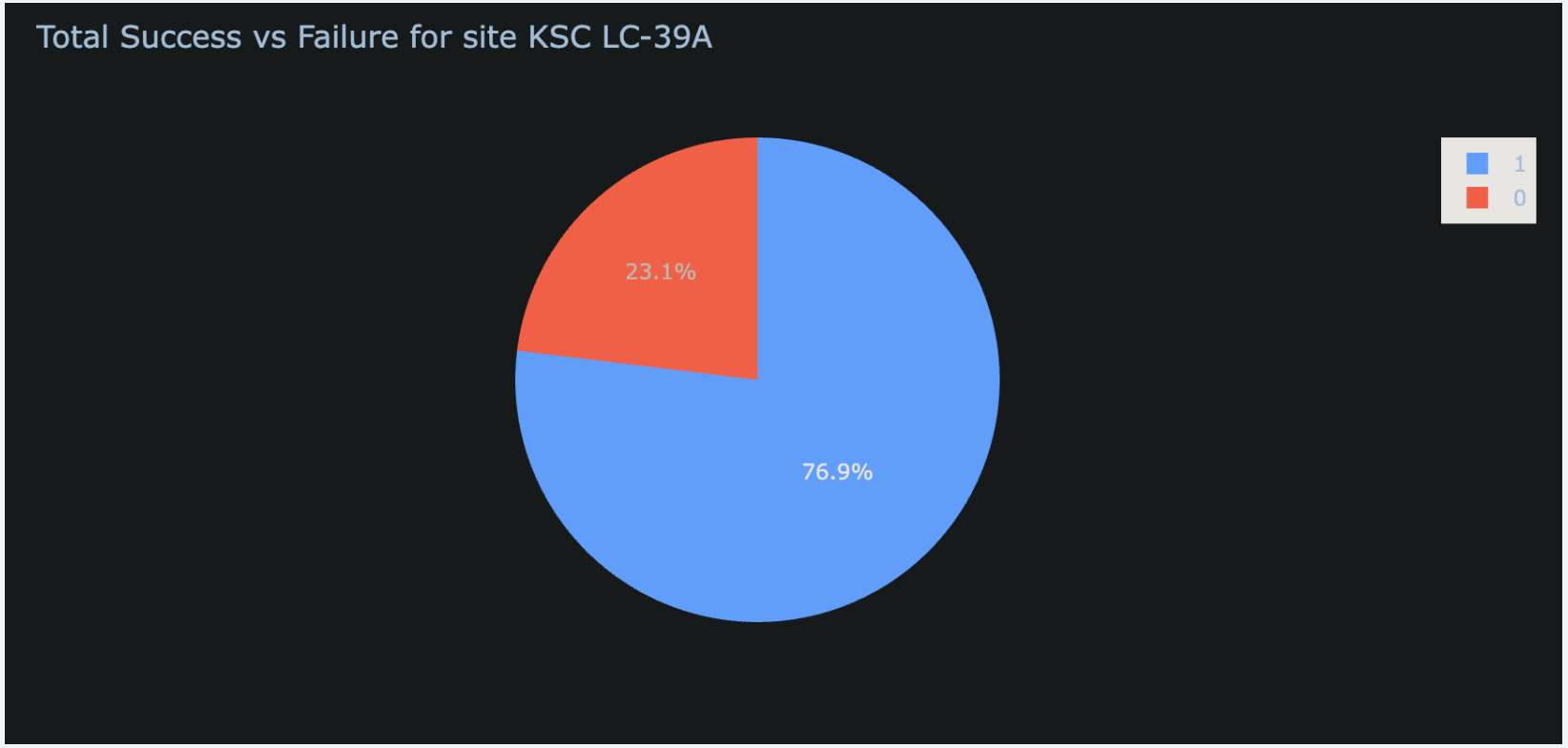
Pie chart showing the relative successful SpaceX launches for each site.

Kennedy Space Center is far ahead of the others at 41.7% successful launches.



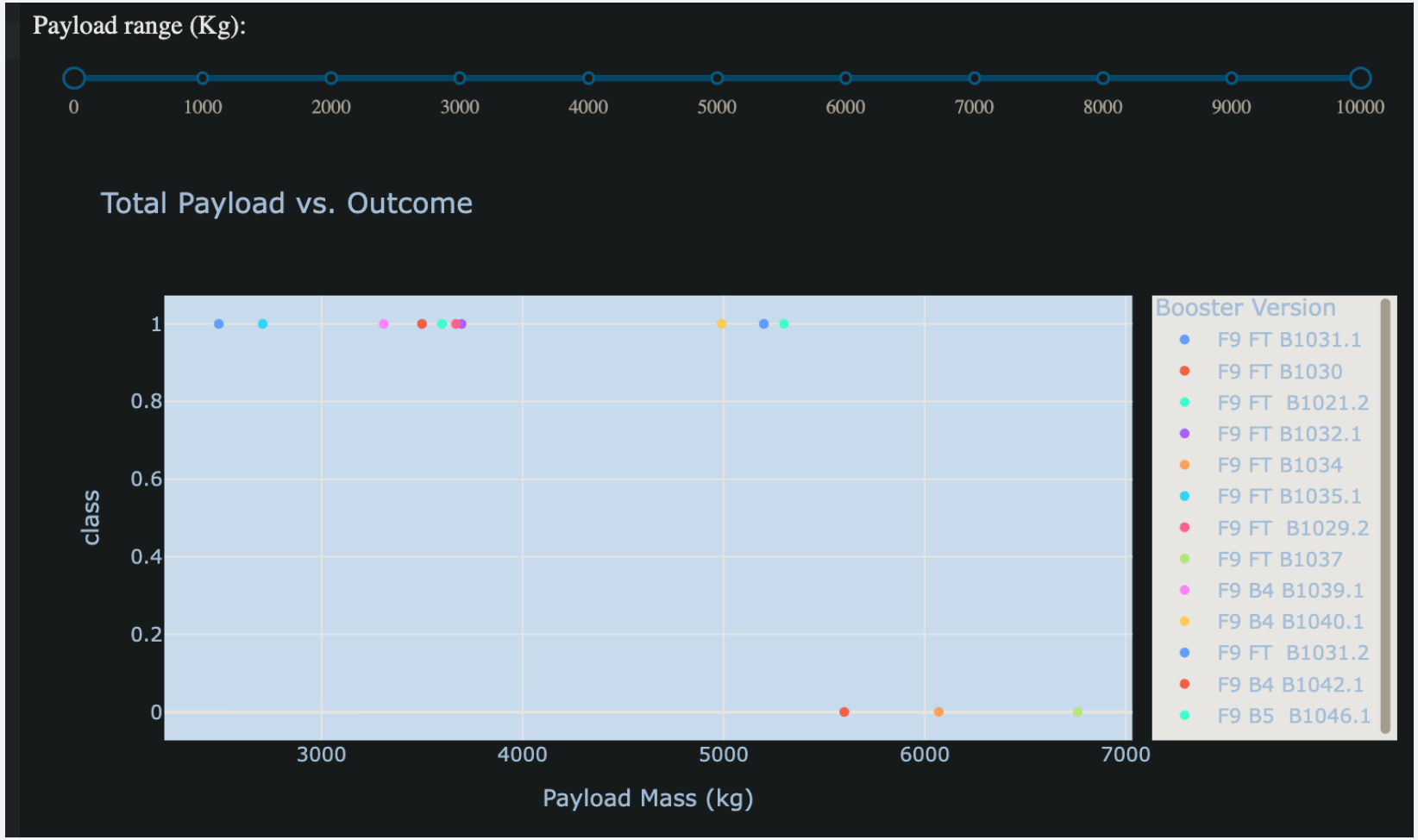
Kennedy Space Center Successes

Pie chart showing that Kennedy Space Center had a 76.9% success rate for SpaceX launches.



Scatter Plot of All Launches by Payload and Booster

A scatter plot showing the launch results for payload mass. Each launch is color coded by the booster version used.

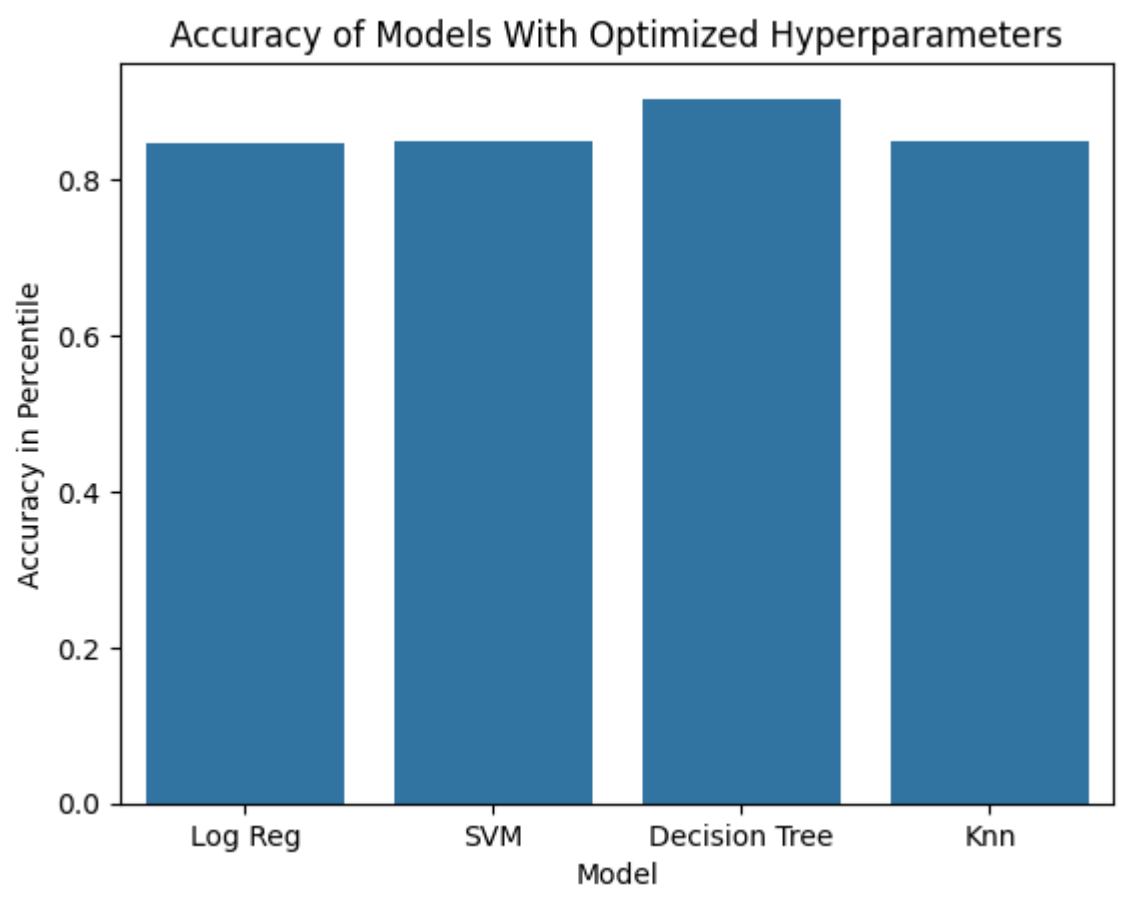


Section 5

Predictive Analysis (Classification)

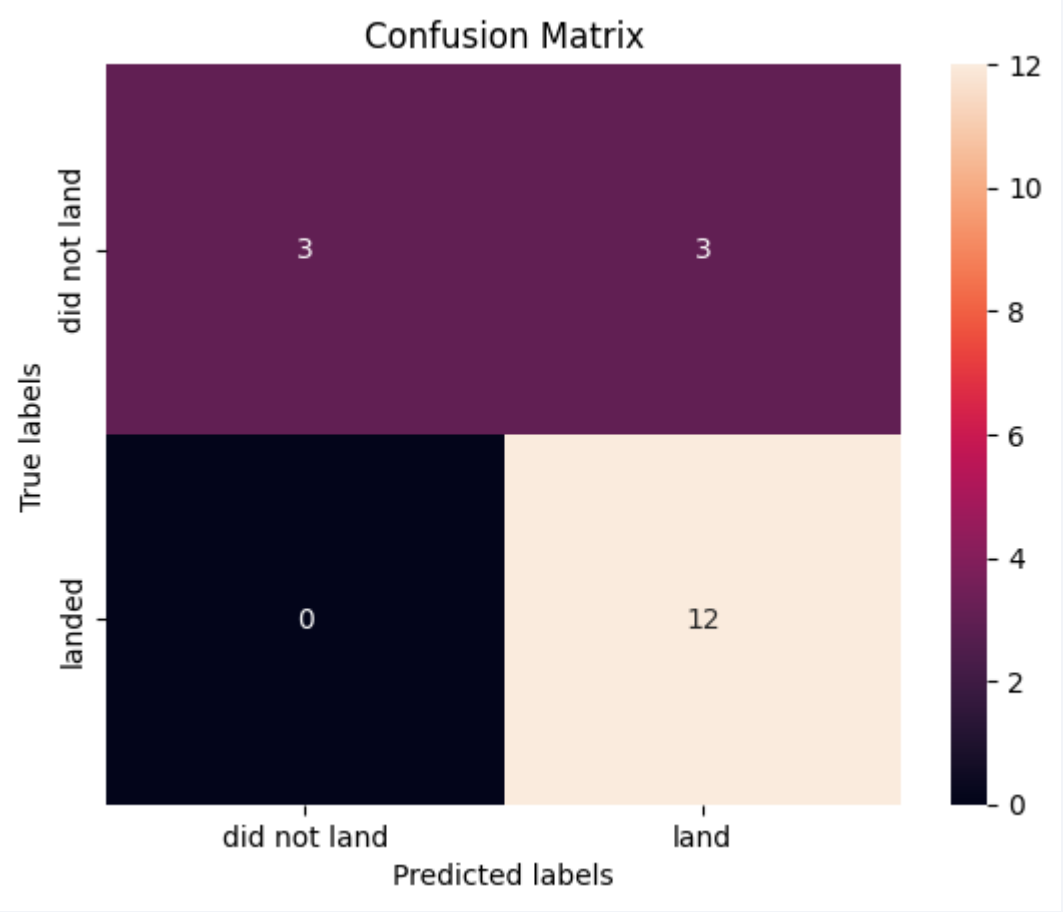
Classification Accuracy

Bar chart showing while all models performed well, the highest possible accuracy model with optimized hyper parameters is the Decision Tree model with an accuracy rating of 87.5%



Confusion Matrix

This confusion matrix of the optimized Decision Tree model shows 3 false positives and 0 negatives. This shows the model is more likely to predict a mission success for a failed mission than predict a mission failure for a successful mission.



Conclusions

- Spacex has made great strides in their effort of creating more cost-efficient space travel by innovating a reusable first stage rocket.
- The VLEO orbit has had the highest success rate among the frequently used orbits.
- Among four predictive models tested, a decision tree classifier had the highest accuracy rate at 87.5%.
- There is potential for further analysis for first stage rocket reusability, and future competitors in this space will need to push themselves to match.

Appendix

Project Github

Thank you!